

# 1. Borealis Preservation Documentation Template

Rev. 2024-12-20

Authors: Mary Catherine Shea, Julie Shi, Robyn Stobbs  
Preservation Expert Group, Digital Research Alliance of Canada

This template is intended to assist Borealis administrators with creating local preservation documentation for your institutional collection. Many aspects of these guidelines should be discussed with stakeholders in your library. When developing these guidelines, you should consider your existing resources and determine if additional commitments are needed.

Per the CTS guidance, “The term preservation plan refers to having a documented approach for defining and implementing preservation actions. The Requirements do not define or differentiate between a preservation policy, plan, strategy, or action plan.” For simplicity, we use “policy” in this template. The guidelines can be adapted to suit your institutional context.

This template is organized into two sections, beginning with higher-level policy suggestions and followed by guidance on preservation practices for research data that may be documented elsewhere, such as in action plans or procedural or workflow documentation.

Documentation based on this template may support your responses to the following 2023-2025 CTS requirements in the [Application Template](#): R0 Background Information, R03 Continuity of Service, R07 Provenance & Authority, R09 Preservation Plan, R11 Workflows, and R13 Reuse.

## 1.1. Important Reminder

All of the institutional policies and procedures for your institutional collection in Borealis should work together in order to make managing your collection easier. Ensuring all of your policies and procedures work together is an important step in their development. Please see this [chart for a visual representation of how institutional Borealis policies and procedures can flow together](#) to cover all the required/needed elements.

<b>1. Borealis Preservation Documentation Template</b>	<b>1</b>
1.1. Important Reminder	1
1.2. Policy: General Sections	2
1.2.1. Issuing Body / Responsible Unit	2
1.2.2. Document Title, Title of Policy	2
1.2.3. Date / Versioning Information	2
1.2.4. Purpose / Objective / Goal	3
1.2.5. Audience	3
1.2.6. Definitions	4

1.2.7. Related Policies	4
1.2.8. References	4
1.3. Policy: Document Specific Sections	4
1.3.1 Scope	5
1.3.2 Principles	5
1.3.3 Roles and Responsibilities	6
1.3.4 Designated Community	6
1.3.5 Levels of Preservation & Storage	7
1.4. Practices: Implementation in the Research Data Context	8
1.4.1 Preservation Activities	8
1.4.2 Ongoing Review and Monitoring	9
1.4.3 Provenance	10
1.4.4 Rights to Preserve – Robyn	11

## 1.2. Policy: General Sections

### 1.2.1. Issuing Body / Responsible Unit

In this section, you may wish to include a logo, and the name of the unit or organization that is responsible for overseeing the guidelines.

**Sample text:** University Libraries, Research Services Division

### 1.2.2. Document Title, Title of Policy

The document should have a title that reflects its objectives and/or document type. For example, if the document is a policy, the word ‘policy’ should be reflected in the title.

**Sample text:** Digital Preservation Policy at the University Libraries for Borealis

### 1.2.3. Date / Versioning Information

It may be helpful to include one or both of the following dates somewhere on your document, such as in the footer: Date Last Updated, Next Review Date.

It may also be helpful to include the version number of the document, especially if this is not the first time such a document has been released at your institution.

If the document is to be considered an official policy, it may also require a policy number. Policy numbers may be provided by your library’s central administrative department based on other policies already in place at your library.

#### 1.2.4. Purpose / Objective / Goal

A purpose, objective, or goal for the document should be included at the beginning. Such a purpose, objective, or goal should state what the document is to be used for, when it should be used, and by whom.

This section may also reference formal commitments to digital preservation, such as in mission statements, mandates, strategic priorities, or other documents at the institutional or repository level.

**Sample text:**

To provide clear guidelines about the preservation of datasets from the University Libraries Institutional Collection in Borealis, and to clearly identify the conditions or circumstances that determine how and when this occurs.

This is in line with our institutional mission to steward the scholarly record into the future.

#### 1.2.5. Audience

If the specific audience for the document is not outlined in the purpose, objective, or goal, or if that audience needs to be defined in more detail, it may be helpful to include a separate audience section.

For a preservation policy, you may need to consider the following in terms of who would be included in your audience:

- Professors (all levels, adjunct, retired, etc.?)
- Librarians (all levels, etc.?)
- Students (undergraduate, graduate, etc.?)
- Classes / Courses / Labs
- Research Projects (institutional, multi-institutional, etc.?)
- Affiliated Organizations (institutions, hospital, research centres, centres of excellence, etc.?)

**Sample text:** This research data preservation policy is intended for all University Libraries for Borealis staff involved in the curation, preservation, and management of research datasets deposited in the Libraries' institutional collection in Borealis. The policy is also of interest to depositors and users of data in this collection or its sub-collections.

For the purposes of this policy, a “depositor” is considered any of the following:

- Faculty and Librarians (for individual research)
- Graduate students (for individual research, including theses and dissertations)

- University Research Projects (including projects at University affiliated organizations and hospitals)

For multi-institutional research projects, please contact [Role] for information on your research data deposit and preservation options.

### 1.2.6. Definitions

A **definition** or **glossary** section should include any specific terms used in the document and their definition as it applies to the document. This is especially useful if the term could be interpreted in multiple ways.

An alternative to having a definition or glossary section within the document could be to have a single document that outlines all Borealis-related definitions used in all the institutional policies and procedures. A link could be used within each document to the single definition or glossary document.

*For the purpose of these templates/guidance documents, a set of definitions will be provided that will cover all documents.*

### 1.2.7. Related Policies

Insert related policies that your preservation policy should be read in conjunction with, as appropriate.

**Sample text:** This policy should be read alongside the University Libraries for Borealis policies on collections, deposit and appraisal, curation, and retention and deaccession.

This policy should also be read alongside the following external documents: the Borealis Preservation Plan.

### 1.2.8. References

Insert relevant references as appropriate. If you've drawn from or adapted any policies and guidance from other libraries or organizations you may wish to reference them as well.

**Sample references:**

- [DPC Digital Preservation Policy Toolkit](#)
- [NASIG Model Digital Preservation Policy](#)
- [NDSA Levels of Digital Preservation](#)
- [OAIS Reference Model](#)
- [Purdue University Research Repository \(PURR\) Digital Preservation Policy](#)
- [Digital Preservation at UVic Libraries](#)

## 1.3. Policy: Document Specific Sections

The following sections refer to specific topics associated with preservation documentation guidelines. Your “preservation guidelines” document/s doesn’t need to include these items by section with or without the titles used, they’re just for reference.

### 1.3.1 Scope

Describe the extent and nature of collection/deposit/content groups that your institution will preserve. Indicate if different preservation levels are offered for different content types (see 1.3.5 Levels of Preservation & Storage). It is recommended to also clearly state what content your institution will not preserve.

Where applicable, refer to institutional policies on collection development, acquisition, appraisal, and deaccessioning.

**Sample text:** The Libraries commits to the long-term preservation of digital content identified as having ongoing cultural or scholarly value. Preservation efforts also apply to the metadata and documentation associated with this content.

Content covered by this policy includes:

- Scholarly outputs
  - Original research data and supporting documentation deposited in Borealis by faculty and graduate students
  - ...
- Library collections, including
  - ...
- Archival holdings, including
  - ...

Content outside of the scope of this policy includes:

- Scholarly outputs
  - Original research data and supporting documentation deposited in Borealis by undergraduate students (with exceptions on a case-by-case basis)
  - Research data generated for replication studies
- ...

### 1.3.2 Principles

Outline the high-level ideas, values, and/or rules that frame your institution’s approach to digital preservation.

Principles may address standards adherence, sustainable practices (technologically, organizationally, financially, and/or environmentally), or minimum preservation activities.

Statements should provide staff with suitable guidance for making preservation decisions and developing action plans and operational procedures.

**Sample text:** The University Libraries for Borealis takes a proactive and risk-aware approach to digital preservation to ensure the ongoing accessibility and usability of research datasets in its care for as long as necessary. Preservation efforts are guided by the following principles:

- The Libraries will prioritize preservation of materials that have enduring value and may preserve these materials at different levels
- The Libraries will preserve content in a manner that is technologically, organizationally, and financially viable

### 1.3.3 Roles and Responsibilities

List the relevant departments/units and staff members primarily responsible for overseeing and implementing aspects of preservation and preservation planning. Also outline the FTE and key responsibilities for each staff member.

**Sample text:**

Department: Library - Research Data Management (RDM)

- Staff members: RDM Librarian (1 FTE), Data Curator (1 FTE)
- Responsibilities: RDM training, appraisal, curation, re-appraisal, policy development, coordinate with campus partners and third-party service providers as appropriate

Department: Administrative Services - Archives and Information Management

- Staff members: Digital Archivist (0.5 FTE)
- Responsibilities: format monitoring, technology watch, re-appraisal, workflow development and implementation

### 1.3.4 Designated Community

Provide a clear definition of the user base that your repository serves. Your Designated Community may be composed of multiple communities, and may change over time. For each community (if multiple are defined), the definition should address:

- The scope and composition of the Community
- Their assumed skillset, knowledge base, and methodologies
- Their expected use cases for materials

Curation and preservation activities by the repository should be guided by the above to ensure the community continues to be able to understand the contents of your repository.

**Sample text:** The University Libraries for Borealis undertakes preservation activities to ensure ongoing access to works of cultural and scholarly significance by its designated community. The Libraries defines its designated community as the University's research community, including faculty and researchers, students, and affiliated research units and organizations across a wide range of disciplines. Ongoing access involves maintaining appropriate infrastructure and

services to enable discovery, access, and use of the Libraries' holdings for learning and research.

Our preservation strategies and activities are informed by the following assumptions about our user community:

- Community at large
  - The community is fluent in either English or French
  - The community has access to an Internet-connected computer and an internet connection that allows for uploading and/or downloading data
- Community members using our research data repository
  - The community is working with formats that can be opened and used using software that is generally available to researchers working within the same domain and the broader Canadian research community
  - The community primarily uses the repository for depositing data associated with publications
- Community members using our institutional repository
  - ...

Our approach to preservation will evolve as our understanding of our community changes.

### 1.3.5 Levels of Preservation & Storage

Not all materials require the same level of preservation. Describe the various levels of preservation and storage that your institution implements. Preservation and storage levels may be defined independently or together.

Where institutions provide two or more levels of preservation and/or storage, you may reference documentation that defines which content is in scope for preservation (see 1.3.1 Scope of this template and the Borealis Institutional Collection Policy Template).

**Levels of Preservation:** Bit-level preservation is generally identified as the minimum level of preservation to ensure that files remain unchanged and integrity is retained over time. Bit-level preservation is achieved by generating a checksum for each file upon capture or ingest, and then ensuring at regular intervals that the checksum of the file in storage continues to match the initial value. This does not ensure that files continue to be accessible, usable, and understandable.

Depending on your repository's capacity, further levels to support ongoing access, use, and understanding may include:

- Normalization: transformation of file content into open formats at time of ingest
- Migration: transformation of file content into contemporary, preferably open formats as normalized formats become obsolete
- Emulation: simulation of the original computing environment to access and interact with the file in its original technical context

Activities that occur across multiple levels, such as processing in a preservation system or auditing, or that relate to the implementation of these levels, such as requesting BagIt exports from the Borealis team, may be described alongside the levels or elsewhere (see 1.4.1 Preservation Activities and 1.4.2 Ongoing Review and Monitoring).

**Levels of Storage:** Describe the levels of storage that your institution is able to provide in support of preservation, addressing the number of copies, locations (and disaster risk profiles, as appropriate), storage solutions, and other criteria.

Common methods for mitigating storage risks is to maintain:

- At least two independent copies of data (i.e. deleting or modifying one copy will not affect the other copy/ies)
- In separate locations (with different risk profiles for more rigorous mitigation)
- On different storage media or systems (e.g. one online and one offline copy)
- Managed by different organizations (especially for vended systems)

**Sample text:** The University Libraries for Borealis commits to provide one or more of the following levels of preservation and storage for its digital holdings, as appropriate for the content type and manageable with available resources:

### **Levels of Preservation**

Level 1: Bit-level - The University Libraries ensure that files retain fixity over time via checksum generation and comparison

- Objective: Users can access copies of the material in its original format with the assurance that its contents remain unchanged

Level 2: Normalization - At the time of capture or ingest, the University Libraries transforms original materials in formats that are closed or not widely supported into preferred formats that are open or well supported at the time of ingest. Bit-level preservation is applied to both original and normalized files.

- Objective: Users can access materials in a format that was open or better supported at the time of ingest and copies retain most or all of the significant properties of the original material

Level 3: Migration - The University Libraries migrates content to current preferred formats as formats selected during normalization become obsolete. Bit-level preservation is applied to both original and normalized files.

- Objective: Users can access materials in a migrated format suitable for the technology available at the time of access, with as many significant properties retained as practically possible. Checksums for migrated copies generated at time of future access match the checksum generated at time of creation

### **Levels of Storage**

Storage levels are based on recommendations in the NDSA Levels of Digital Preservation. Staff maintain documentation about all storage locations, media, and contents:

- Level 1: Two complete, independent copies in separate locations
- Level 2: Three complete copies, with at least one copy in a separate location
- Level 3: Three complete copies, with at least one copy in a location with a different disaster risk profile than that of the other location/s and at least one copy on a different type of storage media

- Level 4: At least three copies in locations, each with a different disaster risk profile, management, and type of storage media

## 1.4. Practices: Implementation in the Research Data Context

The following outlines common processes for implementing preservation policies and programs. This section does not cover all possible preservation activities, instead focusing on key practices noted in the CoreTrustSeal requirements.

### 1.4.1 Preservation Activities

Describe activities that your institution undertakes to achieve the level/s of preservation and/or storage applied to research data. Also indicate who is responsible for each activity (see 1.3.3 Roles and Responsibilities) or if activities are performed by external partners.

Activities completed during appraisal and curation that support or inform preservation efforts may also be referenced. See the Borealis Institutional Collection Curation Framework Template and Borealis Institutional Collection Deposit & Appraisal Guidelines for guidance in these areas.

#### **Sample text:**

University Libraries for Borealis data repository:

- Preservation level: Level 2 - normalization on deposit; Level 3 - migration as needed

Data deposit and curation guidelines encourage depositors to submit files in open, widely recognized formats as much as possible. If files are received in proprietary formats or formats that are not widely supported, the Data Curator will work with the depositor to create a normalized copy in a more accessible format. Both the original and normalized copies are deposited. (See the Curation Framework and Deposit & Appraisal Guidelines for additional activities to ensure robust metadata and documentation.)

After each dataset is published, the Digital Preservation Librarian processes the dataset using our locally-hosted Archivematica system. Datasets are retrieved directly from the repository by Archivematica using an API integration, which retrieves all deposited files and metadata associated with a dataset as well as derivatives generated by the Dataverse software for tabular data (e.g. tab-delimited file, RData file) and metadata (e.g. DDI Codebook, JSON LD, etc.).

Processing in Archivematica is intended to (1) generate a secure and stable copy of each dataset in a standardized digital package, (2) gather the technical metadata necessary to identify and manage our formats and inform plans to migrate to new formats or format versions in future as necessary, and (3) enable migration using Archivematica processes for revising format guidelines and reprocessing packages.

- Storage level: Level 3: Three complete copies, with at least one copy in a location with a different disaster risk profile than that of the other location/s and at least one copy on a different type of storage media

The Ontario Library Research Cloud, hosted by Scholars Portal, is the primary storage location for files uploaded to our data repository. This storage solution distributes three synchronized copies of data to various data centres at Ontario university libraries. We treat this as one independent copy of the data.

Scholars Portal also generates daily backups to tape to protect against data loss. Tape backups are managed using industry-standard tape backup software, with a copy stored offsite for disaster recovery. More details about the repository's preservation practices are outlined in the [Borealis Preservation Plan](#).

For further redundancy, archival packages generated by Archivematica are stored on a local RAID storage array and also written to tape for permanent archiving. As well, the RDM Librarian coordinates with the Borealis service team on an annual basis to receive an independent copy of published datasets as BagIt-compliant packages, which are stored in AWS S3 Glacier Deep Archive storage in the Canada (West) region to account for the most unexpected scenarios.

#### 1.4.2 Ongoing Review and Monitoring

**Community needs:** Explain processes by which relevant departments/units and staff assess and address their Designated Community's ongoing needs. Also, describe methods for fostering communication and feedback to address changes to the Community's access and use requirements. These may include surveys, interviews, and repository links to departmental inboxes.

Staff should also outline how relevant institutional advisory groups or committees formally address these changing needs. If the institution has defined two or more designated communities, any differences in monitoring actions and processes should be described here.

**Technology and formats:** For preservation beyond bit-level, outline processes for reviewing and monitoring the preservation environment for relevant collections, including the risks that technological obsolescence pose to access and use. Institutions should describe actions and workflows for addressing identified technological risks, as well as evolving storage needs and tools to effectively meet digital preservation goals.

**Policies, workflows and documentation:** Institutions should outline their policy review cycle(s) as they relate to digital preservation. Also describe internal practices and workflows to evaluate and refine preservation processes and documentation over time (e.g. evaluating data formats and assessing and updating procedures for refining file normalization procedures for datasets).

**Sample text:**

To monitor **designated community needs**, personnel will:

- Solicit feedback from designated communities via anonymous surveys circulated in the course of biannual research data management group training sessions
- Analyze monthly Borealis usage reports on dataset downloads, file types, and subject categories to garner insights on community user needs
- Determine researcher preservation priorities and needs via staff contributions to and participation in institutional advisory group and committees, national consortia, and professional associations (e.g. Digital Research Alliance of Canada, IASSIST, etc.)

To review and monitor changes in digital preservation **technology and formats**, designated staff will:

- Perform regular format audits, recording formats of stored data and assessing the risk each poses for technological obsolescence
- Propose mitigating actions (e.g. file migration, normalization) on a priorities-based ranking scale
- Periodically assess collections' storage needs as they evolve over time, including determining requirements for geographic location, number of copies, security, and storage service provider for each content or collection type hosted by the institution

**Policies, workflows and documentation:**

Institutional policies, including the digital preservation policy, are reviewed every other year. The committee assesses the previous policy against developments in technology and research infrastructure, changes in related institutional policies and changing community needs, and makes recommendations to ensure the policy continues to align with user needs and the current digital preservation environment.

A sub-group of the digital preservation policy committee, composed of the RDM Librarian and Digital Preservation Librarian, updates or develops related workflows to implement recommended changes. Workflows are tested in consultation with the Data Curator and checked for accuracy and efficacy, before becoming adopted as institutional practice.

### 1.4.3 Provenance

Provenance information is critical for confirming that objects remain authentic and reliable, and changes to data, if any, are authorized and justified. Describe processes for reviewing data and metadata, managing changes, if any, and documenting these changes, including justification and those involved.

**Sample text:** Datasets are reviewed during curation and appraisal, re-appraisal, and audits. Staff maintain detailed records of all dataset reviews and their outcomes, including any changes made, based on standardized checklists and template log files. All logs are stored in a OneDrive folder accessible only to repository staff.

All changes made during appraisal and curation are first discussed with the original depositor. Changes made during reappraisal are reviewed by at least two of three staff among the RDM Librarian, Data Curator, and Digital Preservation Librarian. For all changes during these stages, staff will indicate in the logs:

- the nature of the change
- the reason for the change
- consequences resulting
- the names of those involved
- the date the change was made

Depending on the change required (e.g. format migration), the archival package associated with the dataset may be reprocessed in Archivematica, in which case logs will also be added to the package.

Records and review timelines are otherwise managed in OneDrive, which is subject to regular backups based on the University's backup policy.

A random sample of packages are also retrieved annually from local storage, tape, and AWS to ensure fixity and that files and metadata can still be rendered, used, and understood by the designated community. Audit logs note the date of and person responsible for the audit, whether the audit was successful or not, and a brief description of the packages reviewed and activities conducted.

#### 1.4.4 Rights to Preserve

Repositories need to have the necessary rights and permissions to preserve digital materials. Preservation requires that a repository be able to make copies, migrate and/or transform materials, and modify them if needed to ensure they continue to be accessible and usable in the long term. They also need to know under what terms the items can be made available.

In archives, this transfer of rights is often done through a deed of gift, where the archives becomes the owner of the items and may even also be transferred the copyright. Data repositories tend to deal with rights through terms of use or deposit agreements that specifically grant the repository the necessary rights for preservation, but leave intellectual property rights with the depositor. For example, through the [Borealis Terms of Use](#), users grant Borealis the necessary rights for preservation.

Institutions using Borealis for their institutional repository may additionally have their own terms of use or policy to secure preservation rights allowing the institution to maintain items deposited in their institutional collection, even if they were to change platforms (e.g., migrate off of Borealis), or if they were to use additional platforms or infrastructure to provide preservation. Preservation rights may also be requested through deposit agreements (see the Borealis Institutional Collection Deposit & Appraisal Guidelines for more).

If your institution does not receive rights to preserve (or receives only a limited set of rights), this will limit the preservation activities you are able to do in the future, particularly if the original depositor becomes unreachable. When Creative Commons licences are assigned, the repository can distribute and take preservation actions that are in keeping with the assigned licence. However, if custom terms of use are applied, the repository may not be able to provide access to items, particularly if the designated contact for the dataset stops responding to access requests and is no longer affiliated with your institution.

**Sample text:** Our objective remains the continued access and preservation of deposited datasets for the longer term. To support this objective, the University Libraries reserves the right to make multiple copies for the purposes of security, back up, and preservation. University Libraries may migrate files to any medium, format, or platform for the purposes of preservation, but will not change the content or meaning of the work. Changes made to files for preservation purposes will be recorded in appropriate metadata to ensure authenticity of the files.

This information is also included in section 1.3.7 of the Borealis Institutional Collection Policy Template.

#### 1.4.5 Succession and Continuity Planning

Succession planning involves developing procedures to relocate or transfer materials to another repository in the event of the repository service being decommissioned (e.g., if institutional funding sources ceases, or if the host institution was to close). It can be difficult to make these types of arrangements in advance, but your institution may create sets of principles that will be followed in the event that the service has to be decommissioned.

In the medium term, succession planning may also address continuity of service. For example, the plan may explain what contingencies are in place should a third-party service provider discontinue their service or change the terms of use in such a way that requires your institution to discontinue use of the provider's services.

Rights may also be a concern for the purposes of succession planning, if a repository is being decommissioned and ownership of the items does not fall with the repository, it may be difficult to transfer the items to a new organization unless the existing licence terms are sufficiently open.

**Sample text:** In the event of the termination of our agreement with Borealis, data will be maintained in keeping with our service level agreement by University of Toronto Libraries for six months. During that time we will transfer data to temporary storage operated by the University Libraries while an alternate repository platform and preservation service is set up and/or contracted.

